

Epigenome-wide Association Study of Attention-Deficit/Hyperactivity Disorder Symptoms in Adults

Supplement 1

EWAS Model Equations

The following models were fitted in each cohort:

NTR

$$\begin{aligned} CpGi = & \alpha + \beta_{ADHDSymptoms} \times ADHDSymptoms + \beta_{sex} \times sex + \beta_{age} \times age + \beta_{smoking} \times smoking \\ & + \beta_{gPC1} \times gPC1 + \beta_{gPC2} \times gPC2 + \beta_{gPC3} \times gPC3 + \beta_{monocyte} \times monocyte\% \\ & + \beta_{eosinophil} \times eosinophil\% + \beta_{neutrophil} \times neutrophil\% \\ & + \beta_{arrayrow} \times arrayrow + \beta_{sampleplate1} \times plate1 \dots + \dots \beta_{sampleplate33} \times plate33 \\ & + \varepsilon \end{aligned}$$

Dunedin

$$\begin{aligned} CpGi = & \alpha + \beta_{ADHDSymptoms} \times ADHDSymptoms + \beta_{sex} \times sex + \beta_{smoking} \times smoking \\ & + \beta_{gPC1} \times gPC1 + \beta_{gPC2} \times gPC2 + \beta_{monocyte} \times monocyte\% \\ & + \beta_{eosinophil} \times eosinophil\% + \beta_{neutrophil} \times neutrophil\% + \beta_{mPC1} \times mPC1 \dots \\ & + \dots \beta_{mPC32} \times mPC32 + \varepsilon \end{aligned}$$

E-Risk

$$\begin{aligned} CpGi = & \alpha + \beta_{ADHDSymptoms} \times ADHDSymptoms + \beta_{sex} \times sex + \beta_{smoking} \times smoking \\ & + \beta_{gPC1} \times gPC1 + \beta_{gPC2} \times gPC2 + \beta_{plasmablast} \times plasmablast\% \\ & + \beta_{CD8+CD28-CD45RA-T} \times CD8 + CD28 - CD45RA - T\% + \beta_{CD8T} \times CD8T\% \\ & + \beta_{CD4T} \times CD4T\% + \beta_{NK} \times NK\% + \beta_{monocyte} \times monocyte\% \\ & + \beta_{granulocyte} \times granulocyte\% + \beta_{mPC1} \times mPC1 \dots + \dots \beta_{mPC28} \times mPC28 + \varepsilon \end{aligned}$$

where CpGi is DNA methylation β -value at methylation site *i*, α is the intercept, ADHDSymptoms is ADHD symptoms, sex is coded as 0 for males and 1 for females, age is the age at blood sampling in years (which only varies across individuals in NTR), smoking is smoking status, coded as 0 (never smoked), 1 (former-smoker), 2 (current-smoker) in NTR and Dunedin and coded as 0 (never-smoked) and 1 (current-smoker) in E-Risk, gPC1, gPC2, and gPC3, are PCs 1, 2, and 3 based on genome-wide genotype data, respectively,

monocyte%, eosinophil% and neutrophil%, plasmablast%, CD8+CD28-CD45RA- T%, CD8T%, CD4T%, NK%, and granulocytes%, are percentages of white blood cells, array row is the row of the sample on the Illumina 450k Beadchip (ranging from 1 to 6), plate1...plate33 are 96-wells bisulfite plates (the samples were processed on 34 plates in total), mPC1 ... mPC32, are Principal Components 1 to 32 based on methylation array control probes, and ϵ is residual.

Cohort 1: Netherlands Twin Register

Subjects and samples

The subjects participated in longitudinal survey studies from the Netherlands Twin Register (NTR) (1) and were phenotyped for ADHD symptoms multiple times. Between 2004 and 2011, a subsample was invited to take part in the NTR biobank project (2). The longitudinal survey design and blood sampling procedures have been described in detail previously (1, 2). In the current EWAS, we included individuals for whom the following data were available: phenotype information from the CAARS ADHD index, good quality DNA methylation data, genome-wide SNP data, white blood cell percentages (monocytes, neutrophils and eosinophils), and smoking status, and excluded 'ethnic outliers'; individuals who were outliers on Principal Components (PCs) based on genome-wide genotype data. This left 2258 samples from 2232 subjects (for 26 subjects, data from two longitudinal methylation samples were included). Informed consent was obtained from all participants. The study was approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Centre, Amsterdam, an Institutional Review Board certified by the U.S. Office of Human Research Protections (IRB number IRB00002991 under Federal-wide Assurance FWA00017598; IRB/institute codes, NTR 03-180).

ADHD symptoms

Data on ADHD symptoms were collected in multiple NTR surveys. For the EWAS, data from the CAARS (Conners' Adult ADHD Rating Scales) ADHD index (3), collected in NTR Survey 7 (data collection in 2004; N=1225), Survey 8 (data collection in 2009; N=875), and Survey 10 (data collection in 2013; N=132) were analyzed. The CAARS ADHD index assesses total ADHD symptoms based on 12 items, each of which is scored on a scale from 0 to 4. Thus, the possible range of the CAARS is 0 to 48. For each participant, the ADHD index obtained closest to the date of blood draw was selected. The average time between blood sampling and ADHD symptom assessment was 1.9 years (median=1.5, range: -9.6 years (blood sample

before survey) – 13.1 years (blood sample after survey)). **Supplemental Figure S1a** shows the distribution of the ADHD scores in the 2232 subjects included in the analysis. The CAARS ADHD index is designed to identify adults who are likely to be diagnosed with ADHD. If ADHD index scores are transformed into sex-specific T-scores, a T-score >65 is considered significant in clinical groups (3). The prevalence of ADHD in the NTR based on the T-score cut-off of >65 has been previously described and was 6.8% in women and 7.4% in men (4). Sensitivity analyses were performed on data from the full 30-item CAARS version, included in NTR survey 7, available for 1846 samples. This version includes, in addition to total ADHD symptoms, a 9-item subscale for inattentive symptoms (inattention), and a 9-item subscale for hyperactive-impulsive symptoms (hyperactivity), corresponding to the symptoms that represent the diagnostic criteria of adult ADHD as outlined in DSM-IV-TR. Items from the three subscales did not overlap.

Peripheral blood DNA methylation

DNA methylation was assessed with the Infinium HumanMethylation450 BeadChip Kit (HM450k; Illumina, San Diego, CA, USA) by the Human Genotyping facility (HugeF) of ErasmusMC, The Netherlands (<http://www.glimdna.org/>). Genomic DNA (500 ng) from whole blood was bisulfite-treated using the Zymo EZ DNA Methylation kit (Zymo Research Corp, Irvine, CA, USA), and 4 µl of bisulfite-converted DNA was measured on the Illumina 450k array following the manufacturer's protocol. A number of sample- and probe-level quality checks and sample identity checks were performed (described in detail previously (5)). In short, sample-level QC was performed using MethylAid (6). Probes were set to missing in a sample, if they had an intensity value of exactly zero, a detection $p > .01$, or a bead count of <3. After these steps, probes that failed based on the above criteria in >5% of the samples were excluded from all samples (only probes with a success rate ≥ 0.95 were retained). The methylation data were normalized with functional normalization (7), and normalized intensity values were converted into beta (β)-values.

Covariates

White blood cell percentages were included as covariates in the EWAS to account for variation in cellular composition between whole blood samples, and were obtained as part of the complete blood count (2). The following WBC were included as covariates: monocytes, eosinophils, and neutrophils (lymphocyte percentage was not included because it correlated with neutrophils ($r = -0.9$), and basophil percentage was not included because it showed very little variation between individuals). Information on current and past smoking behavior was collected as part of the NTR biobank project at the moment of blood draw. Smoking status was coded as 0 (never smoked), 1 (former smoker), 2 (current smoker). HM450k array row and

bisulfite plate (dummy-coding) were included as covariates to account for technical variation. The first three PCs based on genome-wide SNP data, which reflect population structure within the Netherlands (8), were included as covariates to account for population differences. Finally, age and sex were included as covariates in NTR, to account for variation in age and sex.

Epigenome-wide association study (EWAS)

EWAS analyses were performed in R software. The association between DNA methylation level and CAARS index was tested with DNA methylation β -value as outcome and predictors CAARS index score, sex, age at blood sampling, smoking status, percentages of monocytes, eosinophils, and neutrophils, HM450k array row, bisulfite plate, and PC1, PC2 and PC3 from genome-wide genotype data. EWAS analyses were performed with generalized estimation equation (GEE) models, which were fitted with the R package 'gee'. The following settings were used: Gaussian link function (for continuous data), 100 iterations, and the 'exchangeable' option to account for the correlation structure within families and within persons.

Cohort 2: Dunedin Multidisciplinary Health and Development Study

Subjects and samples

Participants are members of the Dunedin Multidisciplinary Health and Development Study, a longitudinal investigation of the health and behavior of a representative birth cohort of consecutive births between April 1972 and March 1973 in Dunedin, New Zealand. The cohort of 1,037 children (52% boys) was constituted at age 3 as 91% of eligible births resident in the province. The cohort represents the full range of socioeconomic status on NZ's South Island and matches the NZ National Health and Nutrition Survey on adult health indicators (e.g., BMI, smoking, GP visits) (9). Cohort members are primarily white; approximately 7% self-identify as having any non-white ancestry, matching the South Island. Follow-up assessments were conducted at ages 5, 7, 9, 11, 13, 15, 18, 21, 26, 32, and most recently 38 years, when 95% of the 1,007 living study members underwent assessment in 2010-2012. DNA methylation data were generated for a subset of this cohort (non-Maori participants). In the current EWAS, we included individuals for whom the following data were available: ADHD symptom scores, good quality DNA methylation data, white blood cell counts, smoking status, and genome-wide SNP data, leaving 800 subjects in the analysis sample. Information on current and past smoking behavior was collected on the

same day of blood draw. The study protocol was approved by the institutional ethical review boards of the participating universities. Study members gave informed consent before participating.

ADHD symptoms

ADHD symptoms were assessed at age 38 years, at the same time point as blood collection. Ascertainment of ADHD Symptoms is described in Moffitt *et al.* (2015) (10). Briefly, symptoms were ascertained in 2010-2012 through private structured diagnostic interviews by interviewers with mental-health-related tertiary qualifications and clinical experience. Interviewers received 2 weeks of formal training on the mental health interview, and were trained to inter-rater reliability criterion standard for ascertainment of symptoms, plus re-training continued periodically to prevent drift. Age-38 interviewers were blind to prior data. Because the Dunedin Study's age-38 assessments began in 2010, question/items administered were those reported in leading measures of adult ADHD considered by the DSM-5 working group at that time (11–13). Our interview format followed recommendations from the working group to include behavioral examples relevant for adults (for example, if an item referred to jumping out of seat for children, the interviewer gave examples of difficulty sitting through long meetings or social occasions for adults). The reporting period was the past 12 months. Responses were scored 0=NO/1=YES, and scores summed across the 18 items. To receive an adult ADHD diagnosis, DSM-5 requires that 5 or more inattentive and/or 5 or more hyperactivity-impulsivity symptoms are present. Additional details regarding DSM-5 ADHD symptom scores, how they translate into an ADHD diagnosis, and the prevalence of ADHD (3% in the Dunedin study at age 38 years) have been described previously (10).

Peripheral blood DNA methylation

DNA methylation data were generated using the Infinium HumanMethylation450 BeadChips (Illumina, CA, USA). Whole-blood was collected from the non-Maori participants in K₂EDTA vacutainer tubes (BD, NJ, USA). DNA was extracted from the buffy coat using standard procedures (14, 15). ~500ng of DNA from each sample was treated with sodium bisulfite using the EZ-96 DNA Methylation kit (Zymo Research, CA, USA). Array analysis was performed by the Duke University Molecular Physiology Institute Genomics Core Facility using the iScan platform (Illumina).

Data were processed and normalized using the methylumi (v2.14.0) Bioconductor package from the R statistical programming environment, and subjected to quality control analyses. Briefly, the method corrects Cy3 and Cy5 dye bias and recalculates the betas based on the corrected intensities against a reference array, which defaults to the first chip in the set. Samples were removed if the average detection

p-value was ≥ 0.001 . To confirm genetic identity of the DNA samples, we assessed genotype concordance between SNP probes on the 450K array and data generated using Illumina OmniExpress12v1.1 genotyping BeadChips.

Covariates

Percentages of neutrophils, monocytes, and eosinophils were included as covariates in the EWAS to account for variation in cellular composition between whole blood samples, and were measured using flow cytometry (Sysmex Corporation, Japan) in whole blood samples taken concurrently with the DNA sample. Information on smoking status was obtained at the moment of blood draw and coded as 0 (never smoked), 1 (former smoker), 2 (current smoker). To permit control for technical variation, we used methylation-array control-probe principal components (16). 32 principal components were needed to explain 90% of the variance. These principal components were used as covariates in the analyses. Age was not included as a covariate because blood sample collection and ADHD symptom assessment took place, for all subjects, at the same age (38 years).

Epigenome-wide association study (EWAS)

EWAS analyses were performed in R software. The association between DNA methylation level and ADHD symptoms was tested under a linear model with DNA methylation β -value as outcome and predictors ADHD symptoms, sex, smoking status, neutrophil percentage, monocyte percentage, eosinophil percentage, and PCs1-32 (from genome-wide methylation control probe data), and PCs 1 and 2 from genome-wide SNP data. EWAS models were fitted with the R function `lm()`.

Cohort 3: E-risk Twin Study

Subjects and samples

Participants were members of E-Risk, which tracks the development of a 1994-95 birth cohort of 2,232 British children (17). Briefly, the E-Risk sample was constructed in 1999-2000, when 1,116 families (93% of those eligible) with same-sex 5-year-old twins participated in home-visit assessments. This sample comprised 56% monozygotic (MZ) and 44% dizygotic (DZ) twin pairs; sex was evenly distributed within zygosity (49% male). The study sample represents the full range of socioeconomic conditions in Great Britain, as reflected in the families' distribution on a neighborhood-level socioeconomic index (ACORN [A

Classification of Residential Neighbourhoods], developed by CACI Inc. for commercial use): 25.6% of E-Risk families live in “wealthy achiever” neighborhoods compared to 25.3% nationwide; 5.3% vs. 11.6% live in “urban prosperity” neighborhoods; 29.6% vs. 26.9% in “comfortably off” neighborhoods; 13.4% vs. 13.9% in “moderate means” neighborhoods; and 26.1% vs. 20.7% in “hard-pressed” neighborhoods. E-Risk underrepresents “urban prosperity” neighborhoods because such households are often childless.

Home visits were conducted when participants were aged 5, 7, 10, 12 and most recently, 18 years (93% participation). At age 18, whole blood was collected from 82% (N=1700) of the participants. In the current EWAS, we included individuals for whom the following data were available: ADHD symptom scores, good quality DNA methylation data, smoking status, and genome-wide SNP data leaving 1631 individuals. The Joint South London and Maudsley and the Institute of Psychiatry Research Ethics Committee approved each phase of the study. Study members gave informed consent before participating.

ADHD symptoms

ADHD symptoms were assessed at age 18 years, at the same time point as blood collection. Assessment of ADHD symptoms at age 18 years is described in Agnew-Blais et al (18). Briefly, symptoms were assessed based on private structured interviews with participants regarding 18 symptoms of inattention and hyperactivity-impulsivity according to DSM-5 criteria. Responses were coded 0=NO and 1=YES; and scores were summed across the 18 items. To receive an adult ADHD diagnosis, DSM-5 requires that 5 or more inattentive and/or 5 or more hyperactivity-impulsivity symptoms are present. Additional details regarding DSM-5 ADHD symptom scores, how they translate into ADHD diagnosis, and the prevalence of ADHD (8% in E-Risk at age 18 years) have been described previously (18).

Peripheral blood DNA methylation

DNA methylation data were generated using the Infinium HumanMethylation450 BeadChips run on an Illumina iScan System (Illumina, CA, USA). Whole blood was collected in 10mL K₂EDTA tubes and DNA extracted from the buffy coat using a Flexigene DNA extraction kit (Qiagen, Hilden, Germany) following manufacturer’s instructions.

We assayed 1669 blood samples (out of 1700); 31 samples were not useable (e.g., due to low DNA concentration). ~500ng of DNA from each sample was treated with sodium bisulfite using the EZ-96 DNA Methylation kit (Zymo Research, CA, USA). Twin pairs were randomly assigned to bisulfite-conversion plates and Illumina 450K arrays, with siblings processed in adjacent positions to minimize batch effects.

Fully methylated control samples (CpG Methylated HeLa Genomic DNA; New England BioLabs, MA, USA) were included in a random position on each plate; the distinct DNA methylation profile of this sample enabled us to confirm the experiment was successful and to ensure there were no plate mix-ups or rotations.

Data were imported using the *methyllumIDAT* function in *methyllumi* (19), and subjected to quality control analyses. First, we excluded all samples with median methylated ('M') and unmethylated ('U') intensities <2500. Second, using the ten control probes included on the 450K array, we examined the efficiency of the sodium bisulfite conversion reaction; samples were excluded if their "conversion score" was <80. Third, multidimensional scaling was performed for DNA methylation probes on each of the sex chromosomes and compared to the reported gender. Fourth, to confirm genetic identity of the DNA samples, we assessed genotype concordance between SNP probes on the 450K array and data generated using Illumina OmniExpress24v1.1 genotyping BeadChips.

Samples from 1658 participants passed our QC pipeline. Data were processed with the *pfilter* function from the *wateRmelon* package (Pidsley et al., 2013) excluding 0 samples with >1% of sites with a detection p value >0.05, 567 sites with beadcount <3 in 5% of samples and 1448 probes with >1% of samples with detection p value >0.05. The data were normalized with the *dasen* function from the *wateRmelon* package (20).

Covariates

To control for cell type composition, we used as covariates cell-type proportions estimated from the methylation data (21). Information on smoking status was obtained on the day of blood draw and coded as 0 (never smoked), 1 (current smoker). To permit control for technical variation, we used methylation-array control-probe principal components (16). 28 principal components were needed to explain 90% of the variance. These principal components were used as covariates in the analyses. Age was not included as a covariate because blood sample collection and ADHD symptom assessment took place, for all subjects, at the same age (18 years).

Epigenome-wide association study (EWAS)

EWAS analyses were performed in R software. The association between DNA methylation level and ADHD symptoms was tested with DNA methylation β -value as outcome and predictors ADHD symptoms, sex,

current smoking status, PCs 1-28 (from genome-wide methylation control probe data), estimated cell counts (plasma blasts, CD8+CD28-CD45RA- T cells, naïve CD8 T cells, CD4 T cells, natural killer cells, monocytes, granulocytes), and PCs 1 and 2 from genome-wide SNP data. EWAS analyses were performed with generalized estimation equation (gee) models, which were fitted with the R package 'gee'. The following settings were used: Gaussian link function (for continuous data), 100 iterations, and the 'exchangeable' option to account for the correlation structure within twin pairs.

Motivation for Covariates in the EWAS Analysis

Smoking has widespread effects on DNA methylation (22), and individuals with more ADHD symptoms smoke more (23). We adjusted for smoking status to control for confounding effects of smoking, although we recognize the disadvantage of potentially removing relevant ADHD-related variation. WBC counts were measured in NTR and Dunedin; the following were included as covariates: monocytes, eosinophils, and neutrophils. In E-risk, WBCs were predicted because measured counts were not available; the following were included as covariates: plasma blasts, CD8+CD28-CD45RA- T cells, naïve CD8 T cells, CD4 T cells, natural killer cells, monocytes, granulocytes. Since the molecular contributions to ADHD may vary across individuals with different genetic backgrounds (24), we excluded individuals for whom Principal Components (PCs) based on genome-wide SNPs indicated distinct ancestry (only applicable to NTR), and we included PCs based on genome-wide SNP data as covariates in the EWAS model to account for population structure within each population.

Enrichment Analyses

To study overlap of EWAS signal from the meta-analysis with genetic findings for ADHD, we considered the most recent GWAS for ADHD (25). Since the sample size of the ADHD GWAS is relatively small, and because there appears to be considerable overlap of genetic effects across psychiatric disorders (26), we also considered findings from the much larger GWASs of major depressive disorder (MDD) (27) and schizophrenia (28), and from the GWAS of autism spectrum disorders (ASD) (29). ADHD GWAS summary statistics were obtained from the PGC/iPSYCH consortium. MDD2 GWAS summary statistics (based on the meta-analysis without NTR) were obtained from the PGC. Schizophrenia GWAS summary statistics were obtained from the PGC website (28). ADD summary statistics from the PGC/iPSYCH consortium were

obtained from the iPSYCH website. For each GWAS, we obtained a pruned SNP list, from which we selected all SNPs with a p-value $< 1.0 \times 10^{-4}$ and determined the distance of each Illumina 450k methylation site to each SNP. We considered three windows around GWAS loci, 10 kb, 100 kb, and 1 Mb, to annotate methylation sites, and applied two p-value thresholds to select GWAS loci (genome-wide significant SNPs; $p < 5.0 \times 10^{-8}$ and nominally significant SNPs; $p < 1.0 \times 10^{-4}$). The top 100 CpGs associated with ADHD symptom trajectories in childhood were obtained from Walton *et al.* (30). Because of the small number of epigenome-wide significant loci for the ADHD EWAS, we analysed the top 100 CpGs. Methylation sites associated with schizophrenia were obtained from Hannon *et al.* (31), selecting all epigenome-wide significant CpGs ($p < 1.0 \times 10^{-7}$) from their meta-analysis of all cohorts. Methylation sites previously associated with individual smoking or prenatal exposure to maternal smoking at a false-discovery rate (FDR) $< 5.0\%$ were obtained from recent large meta-analyses by Joehanes *et al.* (22) and Joubert *et al.* (32). To test whether a CpG category was enriched among sites more strongly associated with ADHD symptoms, meta-analysis EWAS test statistics were regressed on each CpG category:

$$|Zscore| = Intercept + \beta_{category\ x} * Category\ x$$

Where $|Zscore|$ represents the absolute Zscore from the EWAS meta-analysis of ADHD symptoms; $\beta_{category\ x}$ represents the estimate for category x , i.e. the change in the EWAS test statistic associated with a one unit change in category x (e.g. being located within 100 kb of SNPs associated with ADHD). For each enrichment test, bootstrap standard errors were computed with 2000 bootstraps with the R-package 'simpleboot'.

Statistical significance was assessed after Bonferroni correction for the total number of GWASs of psychiatric traits (N=4) and EWASs of traits and exposures (N=4); $\alpha = 0.05/8 = 0.006$.

Differentially Methylated Regions

We used the python module 'Comb-p'(33) to scan for regions in which multiple correlated methylation sites showed evidence for association with ADHD symptoms. Comb-p corrects EWAS p-values for the correlation between sites within a particular window and calculates an overall p-value for the region. To consider regions of variable sizes, we applied a step-wise approach, starting with correlated CpGs within an arbitrarily chosen window of 50 bp (seed P-value < 0.01 , minimum of 2 probes) and then stepwise increasing the window size up till 1500 (or until no more significant regions were detected). In this, we

considered the window sizes 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, and 1500 bp. Šidák correction, as implemented in Comb-p, was applied to calculate the p-values for DMRs. For each DMR, Šidák correction accounts for the number of possible tests, defined as the total bases covered by all input probes divided by the size of the region. We report significant regions (Šidák $p < 0.05$) with at least two methylation sites within a 500 bp window. Comb-p was applied to the EWAS results from each of the three cohorts, separately, and on the EWAS meta-analysis. DMR analysis may detect loci where multiple CpGs are associated with ADHD symptoms, but where the individual associations are not significant in EWAS due to insufficient power. Regional plots of DMRs were created with coMET (34).

Gene Expression in Cis

We examined if DNA methylation is associated with gene expression levels *in cis*, using an independent whole blood RNA-sequencing dataset (which did not include NTR participants) from the Biobank-based Integrative Omics Study (BIOS) consortium that was described previously (35). This study tested associations between genome-wide CpGs and transcripts *in cis* (<250 kb). In short, methylation and expression levels in whole-blood samples ($n=2,101$) were quantified with Illumina Infinium HumanMethylation450 BeadChip arrays and with RNA-seq (2x50bp paired-end, Hiseq2000, >15M read pairs per sample). For each target CpG (sites with ADHD symptom EWAS meta-analysis $p < 1 \times 10^{-5}$ or located in significant DMRs), we identified transcripts *in cis* (<250 kb), for which methylation levels were significantly associated with gene expression levels at the experiment-wide threshold applied by this study (FDR<5.0%), after regressing out mQTL and eQTL effects. None of the samples included in the EWAS of ADHD symptoms was included in the expression analysis.

mQTLs

Previously published data from the BIOS consortium were used to look up if top-DMPs and CpGs in DMRs were significantly associated with mQTLs in blood, at the experiment-wide threshold applied by this study (FDR<5.0%, (35)). This study tested both *cis* and *trans* mQTL relationships and was performed on 3841 peripheral blood samples (including a subset of the samples from NTR).

Correlation Between DNA Methylation Level in Blood and Brain

Previously published correlations between DNA methylation levels in blood and four brain regions from matched samples (prefrontal cortex, entorhinal cortex, superior temporal gyrus and cerebellum) were obtained from Hannon *et al* (36). Statistical significance was assessed after Bonferroni correction for the number of brain regions (=4) multiplied by the number of DMPs (=3) plus the number of DMRs tested (=25): $4 * 28=112$, giving an alpha of $0.05/112=4.46 \times 10^{-4}$. The online tool was used to plot DNA methylation levels in blood and brain (<https://epigenetics.essex.ac.uk/bloodbrain/>).

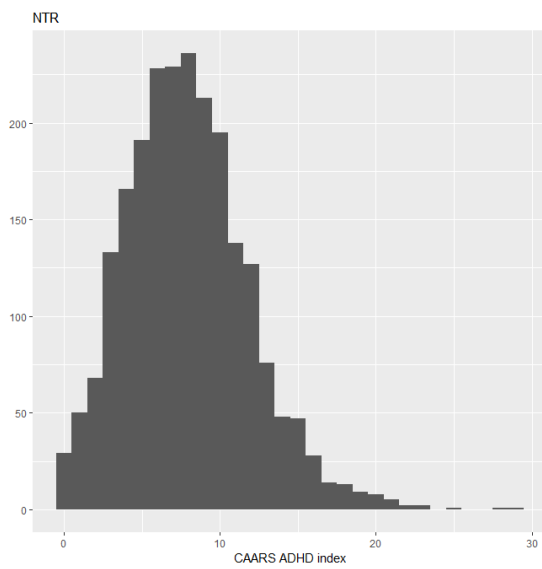
Power Analysis

Power analysis was performed to estimate the sample size required to achieve 80% power to detect the associations between methylation and ADHD symptoms for the three top-DMPs from the meta-analysis at an alpha of 1.0×10^{-7} (epigenome-wide significance), given their effect size observed in NTR. Power analysis was performed with the function `pwr.f2.test()` from the R-package `pwr`. The percentage of variance explained (r^2) at each of the three ADHD top sites was obtained by squaring the correlation for CpGi (r_i), which was derived as follows:

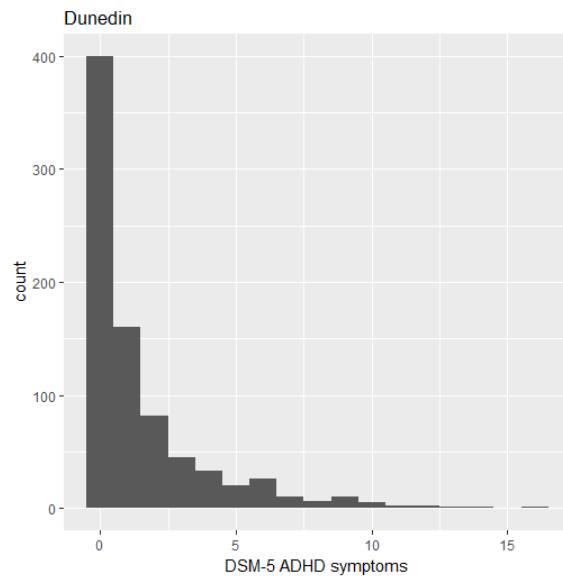
$$r_i = \frac{\beta_i}{\left(\frac{sdy_i}{sdx}\right)}$$

where β_i is the beta from the regression of methylation level on ADHD symptoms in NTR for CpGi, sdy_i is the standard deviation of DNA methylation level residuals of CpGi (obtained in NTR after adjusting methylation levels for covariates), and sdx is the standard deviation of ADHD symptoms in NTR.

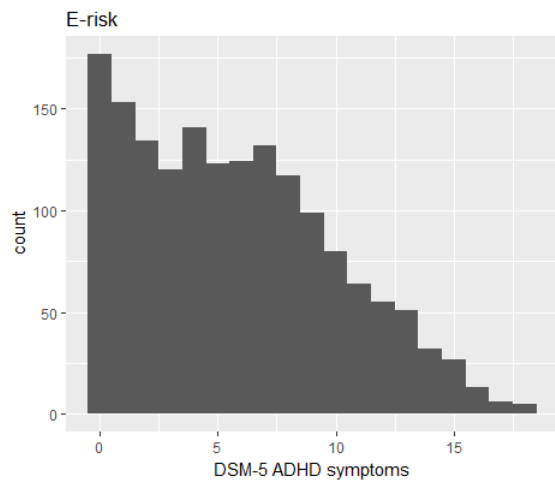
a)



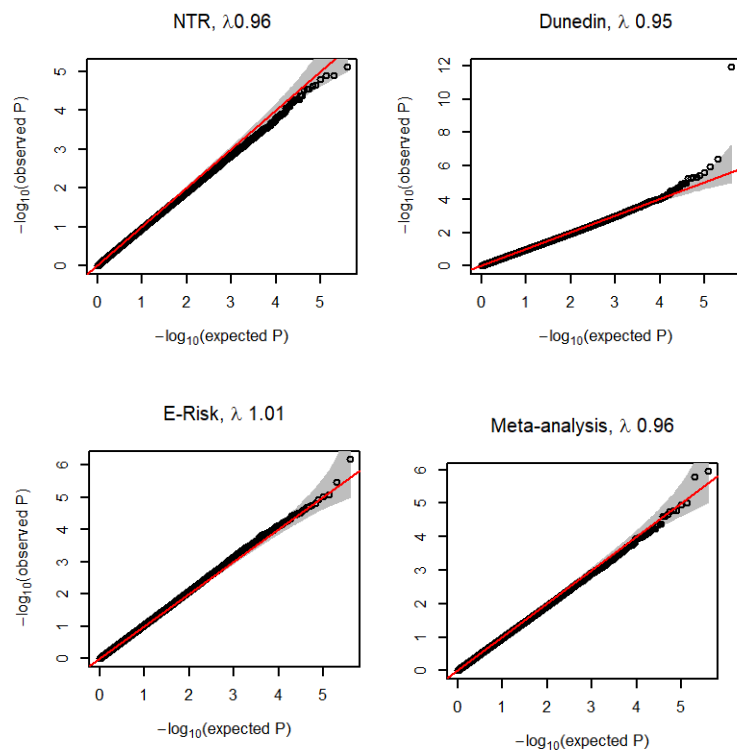
b)



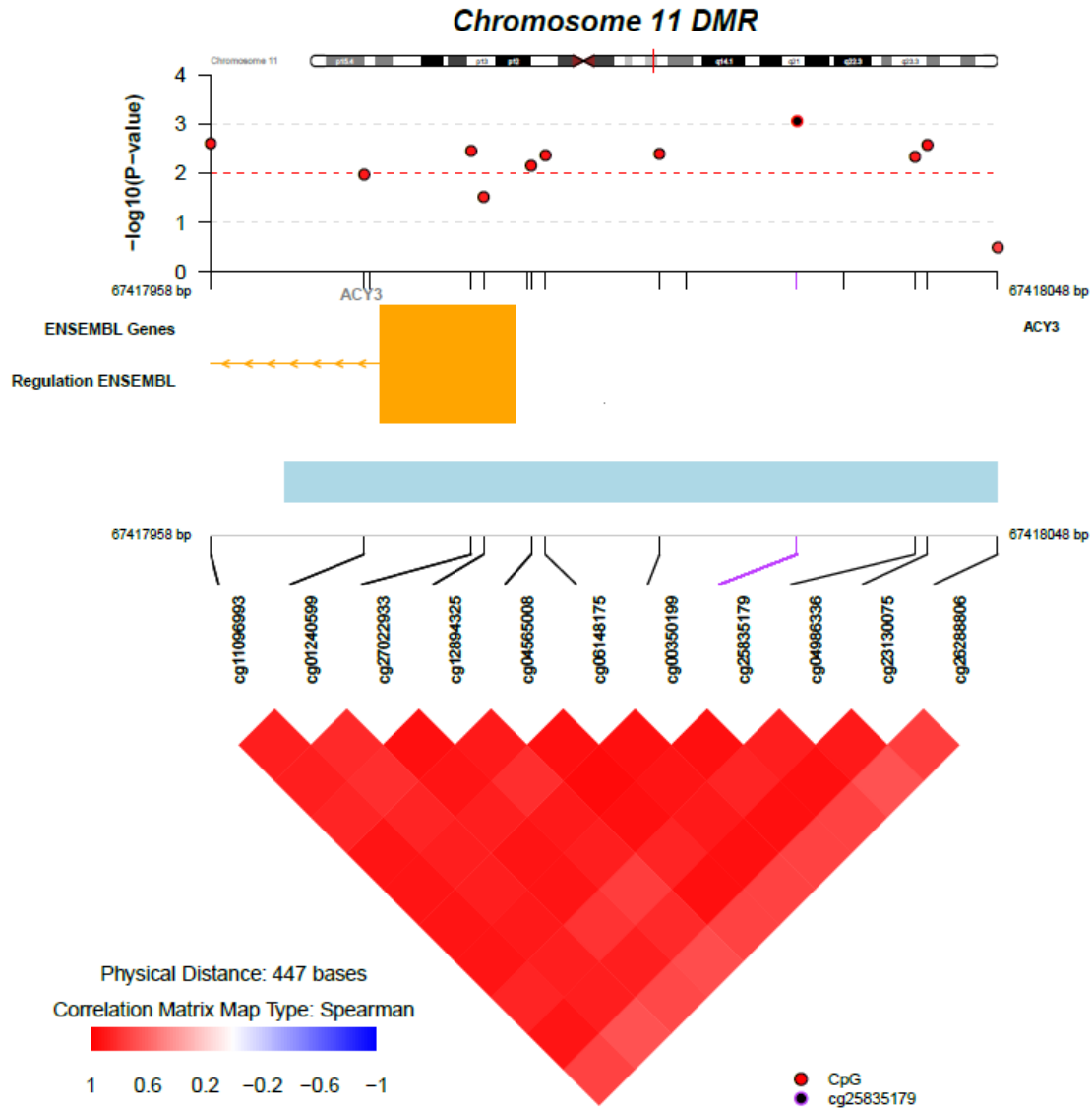
c)



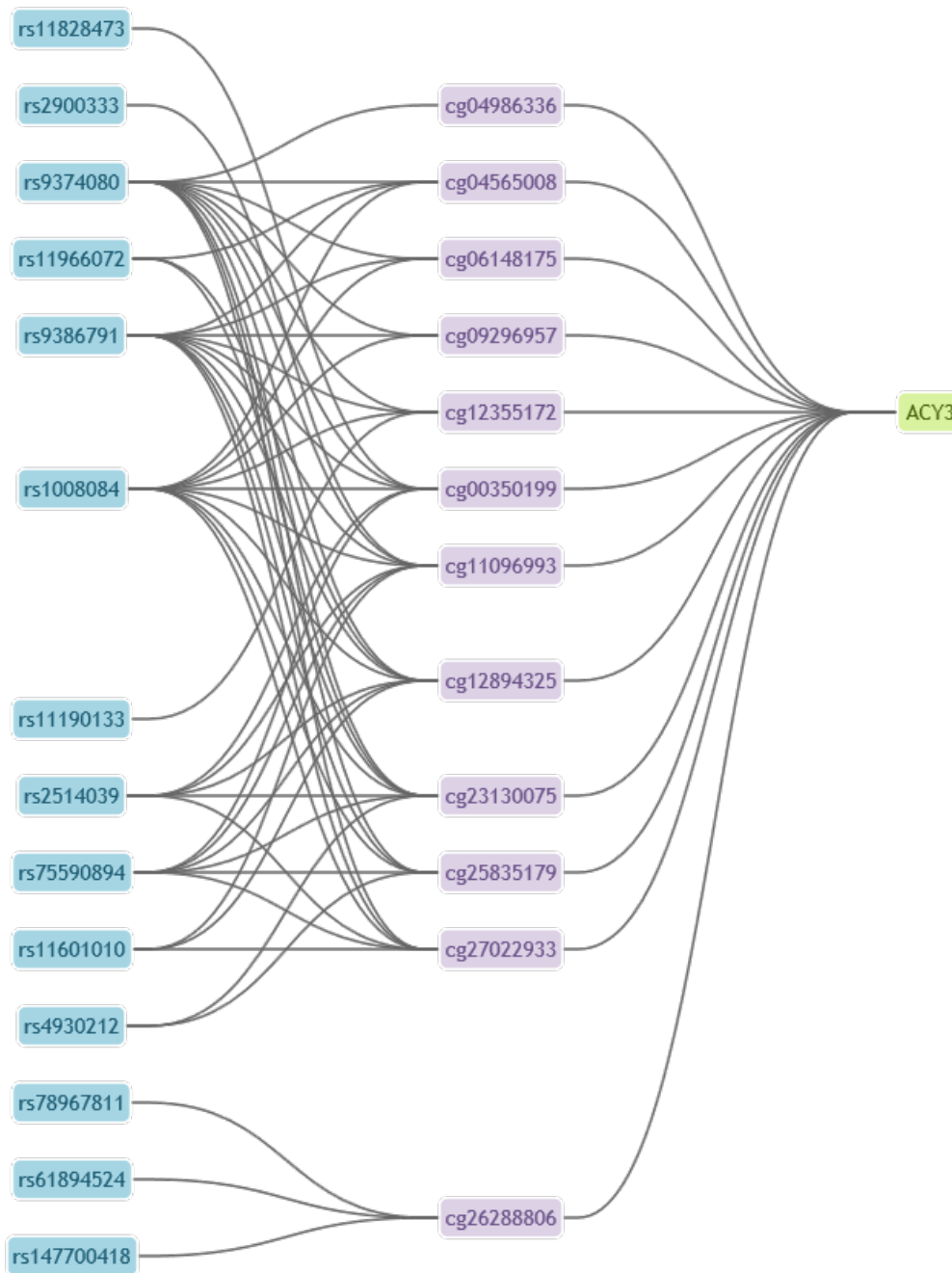
Supplemental Figure S1. Histograms of ADHD symptoms: a) Netherlands Twin Register, b) Dunedin Study, c) E-risk Study



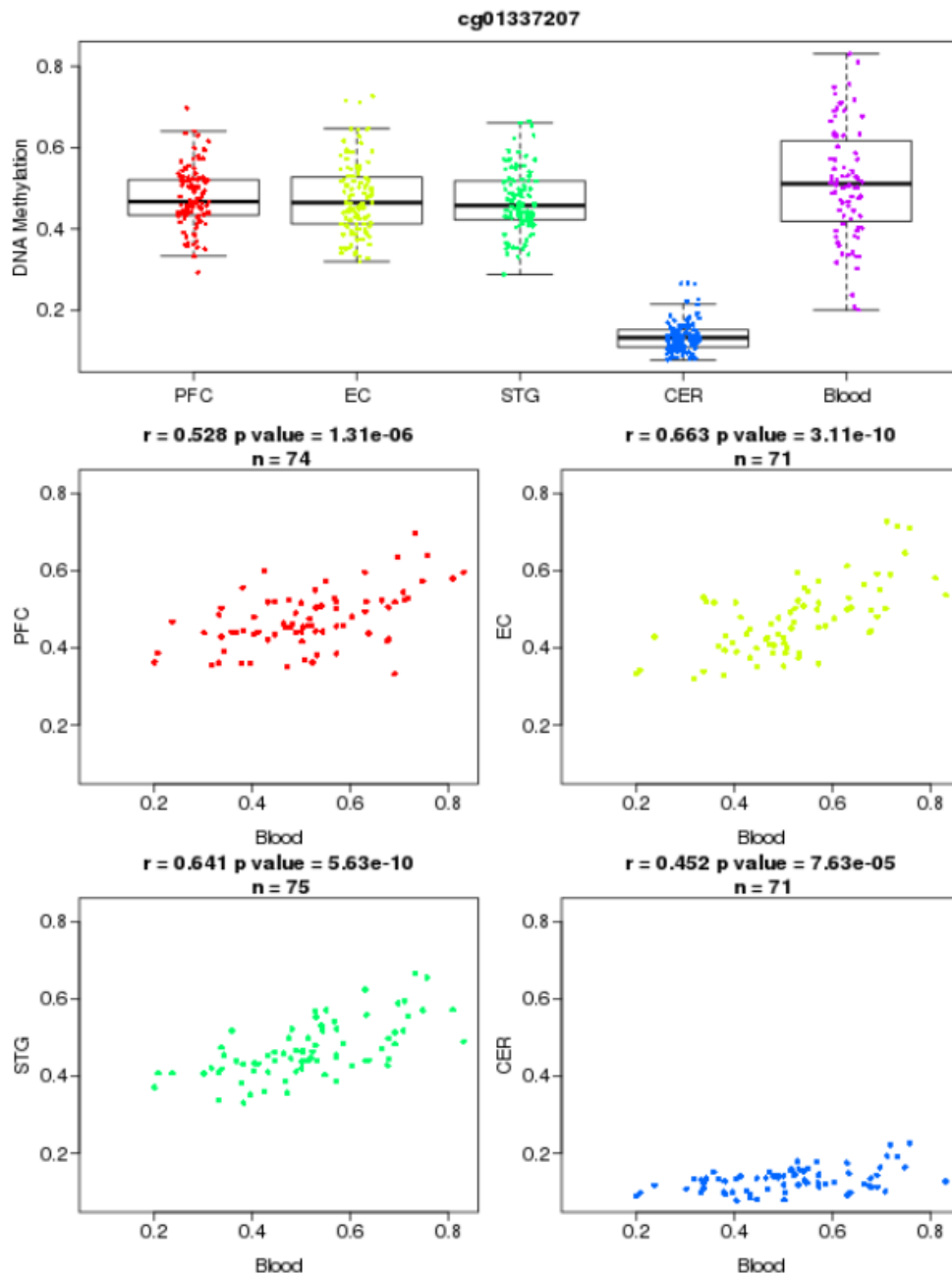
Supplemental Figure S2. QQ plots.



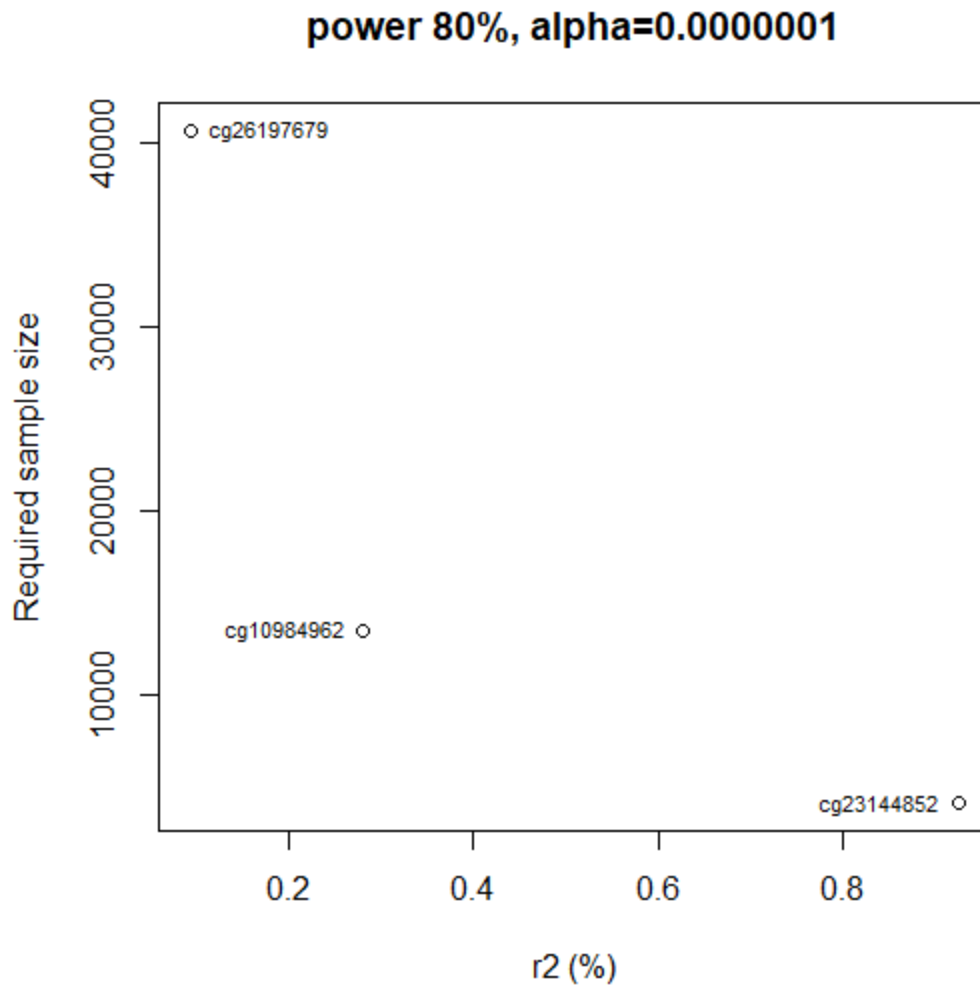
Supplemental Figure S3. Regional plot of a DMR on chromosome 11 that is significantly associated with ADHD symptoms in NTR and with *ACY3* RNA levels in the BIOS consortium (higher methylation levels correlate with lower transcript levels and with lower ADHD symptoms). The top panel shows the EWAS results for ADHD symptoms in NTR. P-values for the individual CpG sites are plotted, with the most strongly associated CpG highlighted in blue. Overall, the region is significantly associated with ADHD symptoms ($p= 7.6 \times 10^{-9}$). The middle panel shows the *ACY3* gene, of which RNA levels are significantly associated with the methylation level of CpG sites in this DMR, and several regulatory tracks. The Regulatory elements track from ENSEMBL shows a promoter-associated (blue) element that overlaps with this DMR. The bottom panel shows the correlations between methylation levels of CpGs in the DMR in NTR, illustrating strong positive correlations between CpGs. The figure was created with coMET (34).



Supplemental Figure S4. Network of mQTLs and methylation sites for a DMR on chromosome 11 that is significantly associated with ADHD symptoms in NTR and with *ACY3* RNA levels in the BIOS consortium. Trans-mQTL SNPs include rs11966072, rs9386791, rs9374080, and rs1008084 (chromosome 6), rs11190133 (chromosome 10), and rs2900333 (chromosome 12). All other SNPs are located in *cis*. The network results (SNP-CpG and CpG-RNA relationships) are based on BIOS consortium data.



Supplemental Figure S5. Example of a CpG site located in an ADHD-associated DMR on chromosome 6 with correlated DNA methylation levels in blood and brain. PFC=prefrontal cortex, EC= entorhinal cortex, STG=superior temporal gyrus, CER=cerebellum.



Supplemental Figure S6. Power analysis results for top-DMPs. The figure shows the required sample size (y-axis) to detect CpGs at an alpha of 1.0×10^{-7} , based on their effect size (r^2 : proportion of variance explained) in NTR (x-axis). The power analysis was performed for the three top sites from the meta-analysis. Note that the true effect size and power will be different in the case of heterogeneity across cohorts.

Supplemental References

1. Willemsen G, Vink JM, Abdellaoui A, den Braber A, van Beek JHDA, Draisma HHM, *et al.* (2013): The Adult Netherlands Twin Register: twenty-five years of survey and biological data collection. *Twin Res Hum Genet.* 16: 271–81.
2. Willemsen G, de Geus EJ, Bartels M, van Beijsterveldt CE, Brooks a I, Estourgie-van Burk GF, *et al.* (2010): The Netherlands Twin Register biobank: a resource for genetic epidemiological studies. *Twin Res Hum Genet.* 13: 231–245.
3. Conners, CK; Erhardt, D; Sparrow E (1999): *Conners' Adult ADHD Rating Scales (CAARS) technical manual.* North Tonawanda, NY; Multi-Health Systems.
4. Boomsma DI, Saviouk V, Hottenga JJ, Distel MA, de Moor MHM, Vink JM, *et al.* (2010): Genetic epidemiology of attention deficit hyperactivity disorder (ADHD index) in adults. *PLoS One.* 5. doi: 10.1371/journal.pone.0010621.
5. van Dongen J, Nivard MG, Willemsen G, Hottenga J-J, Helmer Q, Dolan C V, *et al.* (2016): Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat Commun.* 7: 11115.
6. Iterson M Van, Tobi EW, Slieker RC, Hollander W Den, Slagboom PE, Heijmans BT (2014): MethylAid : visual and interactive quality control of large Illumina 450k datasets. *Bioinformatics.* 30: 3435–3437.
7. Fortin J, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, *et al.* (2014): Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol.* 1–17.
8. Abdellaoui A, Hottenga J, Knijff P De, Nivard MG, Xiao X, Scheet P, *et al.* (2013): Population structure , migration , and diversifying selection in the Netherlands. *Eur J Hum Genet.* 21: 1277–1285.
9. Poulton R, Moffitt TE, Silva PA (2015): The Dunedin Multidisciplinary Health and Development Study: overview of the first 40 years, with an eye to the future. *Soc Psychiatry Psychiatr Epidemiol.* 50.
10. Moffitt TE, Houts R, Asherson P, Belsky DW, Corcoran DL, Hammerle M, *et al.* (2015): Is adult ADHD a childhood-onset neurodevelopmental disorder? Evidence from a four-decade longitudinal cohort study. *Am J Psychiatry.* 172: 967–977.
11. Barkley RA, Murphy KR, Fischer M (2008): *ADHD in adults: What the science says.* *ADHD adults What*

- Sci says*. New York: Guilford Press. doi: 10.1176/appi.ps.59.9.1070.
12. Kessler RC, Green JG, Adler LA, Barkley RA, Chatterji S, Faraone S V., *et al.* (2010): Structure and diagnosis of adult attention-deficit/hyperactivity disorder: Analysis of expanded symptom criteria from the adult ADHD clinical diagnostic scale. *Arch Gen Psychiatry*. 67: 1168–1178.
 13. Manuzza S (2008): *Diagnosing ADHD in adults: DSM-IV controversies and DSM-V recommendations. Review prepared for the American Psychiatric Association.* (D.-V. D. B. D. Workgroup, editor). Washington DC: American Psychiatric Association.
 14. Bowtell DDL (1987): Rapid isolation of eukaryotic DNA. *Anal Biochem*. 162: 463–465.
 15. Jeanpierre M (1987): A rapid method for the purification of DNA from blood. *Nucleic Acids Res*. 15: 9611.
 16. Lehne B, Drong AW, Loh M, Zhang W, Scott WR, Tan S-T, *et al.* (2015): A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol*. 16: 37.
 17. Moffitt TE, Adlam A, Affleck G, Andreou P, Aqan-Assee J, Arseneault L, *et al.* (2002): Teen-aged mothers in contemporary Britain. *J Child Psychol Psychiatry Allied Discip*. 43: 727–742.
 18. Agnew-Blais JC, Polanczyk G V., Danese A, Wertz J, Moffitt TE, Arseneault L (2016): Evaluation of the persistence, remission, and emergence of Attention-deficit/hyperactivity disorder in young adulthood. *JAMA Psychiatry*. 73: 713–720.
 19. Davis S, Du P, Bilke S, Triche J, Bootwalla M (2015): methylumi: Handle Illumina methylation data. R package version 2.14.0. .
 20. Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC (2013): A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics*. 14: 293.
 21. Houseman E, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, *et al.* (2012): DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 13: 86.
 22. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, *et al.* (2016): Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet*. 9: 436–447.

23. Elkins IJ, Saunders GRB, Malone SM, Keyes MA, Samek DR, McGue M, Iacono WG (2018): Increased risk of smoking in female adolescents who had childhood ADHD. *Am J Psychiatry*. 175: 63–70.
24. Cuffe SP, Moore CG, McKeown RE (2005): Prevalence and correlates of ADHD symptoms in the national health interview survey. *J Atten Disord*. 9: 392–401.
25. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, *et al.* (2017): Discovery Of The First Genome-Wide Significant Risk Loci For ADHD. *bioRxiv*. 145581.
26. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, *et al.* (2015): An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 47: 1236–1241.
27. Wray NR, Sullivan PF (2017): Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *bioRxiv*. . doi: <https://doi.org/10.1101/167577>.
28. Ripke S, Neale BM, Corvin A, Walters JTR, Farh K-H, Holmans PA, *et al.* (2014): Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 511: 421–427.
29. Grove J, Ripke S, Als TD, Mattheisen M, Walters R, Won H (2017): Common risk variants identified in autism spectrum disorder. *bioRxiv*. . doi: <http://dx.doi.org/10.1101/224774>doi.
30. Walton E, Pingault J, Cecil CAM, Gaunt TR, Relton CL, Mill J, Barker ED (2016): Epigenetic profiling of ADHD symptoms trajectories : a prospective , methylome-wide study. 22: 250–256.
31. Hannon E, Dempster E, Viana J, Burrage J, Smith AR, Macdonald R, *et al.* (2016): An integrated genetic-epigenetic analysis of schizophrenia: evidence for co-localization of genetic associations and differential DNA methylation. *Genome Biol*. 17: 176.
32. Joubert BR, Felix JF, Yousefi P, Bakulski KM, Just AC, Breton C, *et al.* (2016): DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *Am J Hum Genet*. 98: 680–696.
33. Pedersen BS, Schwartz DA, Yang I V, Kechris KJ (2012): Comb-p : software for combining , analyzing , grouping and correcting spatially correlated P -values. 28: 2986–2988.
34. Martin TC, Yet I, Tsai PC, Bell JT (2015): coMET: Visualisation of regional epigenome-wide association scan results and DNA co-methylation patterns. *BMC Bioinformatics*. 16. doi: 10.1186/s12859-015-0568-2.

35. Bonder MJ, Luijk R, Zhernakova D V., Moed M, Deelen P, Vermaat M, *et al.* (2017): Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet.* 49: 131–138.
36. Hannon E, Lunnon K, Schalkwyk L, Mill J (2015): Interindividual methylomic variation across blood, cortex, and cerebellum: Implications for epigenetic studies of neurological and neuropsychiatric phenotypes. *Epigenetics.* 10: 1024–1032.